1. In a linear equation, what is the difference between a dependent variable and an independent variable?

Ans:  Algebraically, a linear equation typically takes the form y = mx + b, where m and b are constants, x is the independent variable, y is the dependent variable. The slope tells us how the dependent variable (y) changes for every one unit increase in the independent (x) variable, on average. The variables in a study of a cause-and-effect relationship are called the independent and dependent variables. The independent variable is the cause. Its value is independent of other variables in your study. The dependent variable is the effect

1. What is the concept of simple linear regression? Give a specific example.

Ans:In simple linear regression when we have a single input, we can use statistics to estimate the coefficients.consider independent feature is the experience i.e X and the respective salary Y is the dependent variable. Let’s assume there is a linear relationship between X and Y then the salary can be predicted using:

A screenshot of a computer

Description automatically generated

θ1: intercept

θ2: coefficient of x

Once we find the best θ1 and θ2 values, we get the best-fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

For example, suppose that height was the only determinant of body weight.In this example, if an individual was 70 inches tall, we would predict his weight to be: Weight = 80 + 2 x (70) = 220 lbs. In this simple linear regression, we are examining the impact of one independent variable on the outcome.

3. In a linear regression, define the slope.

Ans:  A slope of a line is the change in y coordinate with respect to the change in x coordinate.In linear regression it indicates change in dependent variable as function of change in independent variable.

4. Determine the graph's slope, where the lower point on the line is represented as (3, 2) and the higher point is represented as (2, 2).

Ans: slope is =delY/Del x =0/1=0

5. In linear regression, what are the conditions for a positive slope?

Ans:  If the slope is positive, y increases as x increases, and the function runs "uphill" (going left to right). If the slope is zero, y does not change, thus is constant—a horizontal line.

6. In linear regression, what are the conditions for a negative slope?

Ans:  If the slope is negative, y decreases as x increases and the function runs downhill.

7. What is multiple linear regression and how does it work?

Ans: Multiple linear regression refers to a statistical technique that uses two or more independent variables to predict the outcome of a dependent variable. The technique enables analysts to determine the variation of the model and the relative contribution of each independent variable in the total variance. The formula for a multiple linear regression is:

y = {\beta_0} + {\beta_1{X_1}} + … + {{\beta_n{X_n}} + {\epsilon}

* y = the predicted value of the dependent variable
* B_0 = the y-intercept (value of y when all other parameters are set to 0)
* B_1X_1 = the regression coefficient (B_1) of the first independent variable (X_1) (a.k.a. the effect that increasing the value of the independent variable has on the predicted y value)
* … = do the same for however many independent variables you are testing
* B_nX_n = the regression coefficient of the last independent variable
* \epsilon = model error (a.k.a. how much variation there is in our estimate of y)

To find the best-fit line for each independent variable, multiple linear regression calculates three things:

* The regression coefficients that lead to the smallest overall model error.
* The t statistic of the overall model.
* The associated [p value](https://www.scribbr.com/statistics/p-value/) (how likely it is that the t statistic would have occurred by chance if the [null hypothesis](https://www.scribbr.com/statistics/null-and-alternative-hypotheses/#definition) of no relationship between the independent and dependent variables was true).

It then calculates the t statistic and p value for each regression coefficient in the model.

8. In multiple linear regression, define the number of squares due to error.

Ans: The mean squared error (MSE) tells how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the “errors”) and squaring them. It's called the mean squared error (MSE) as we are finding the average of a set of errors.

9. In multiple linear regression, define the number of squares due to regression.

Ans: Sum of squares is a statistical technique used in regression analysis to determine the dispersion of data points. In a regression analysis, the goal is to determine how well a data series can be fitted to a function that might help to explain how the data series was generated.

10.In a regression equation, what is multicollinearity?

Ans: Multicollinearity occurs when two or more independent variables are highly correlated with one another in a regression model which indicates that an independent variable can be predicted from another independent variable in a regression model.

11. What is heteroskedasticity, and what does it mean?

Ans: Heteroskedasticity (also spelled heteroscedasticity) refers to the error variance, or dependence of scattering, within a minimum of one independent variable within a particular sample. A common cause of variances outside the minimum requirement is often attributed to issues of data quality. [Heteroskedasticity is a **statistical term that describes data with unequal variability across different values of an independent variable or over time**](https://www.bing.com/ck/a?!&&p=cccf0d6740305d4bJmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc5NA&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+heteroskedasticity%2c+and+what+does+it+mean%3f&u=a1aHR0cHM6Ly93d3cuc3RhdGlzdGljc2hvd3RvLmNvbS9oZXRlcm9zY2VkYXN0aWNpdHktc2ltcGxlLWRlZmluaXRpb24tZXhhbXBsZXMv&ntb=1). [It means that the data does not have a constant or uniform spread or dispersion on a scatter plot](https://www.bing.com/ck/a?!&&p=fe300249845f02b3JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc5Nw&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+heteroskedasticity%2c+and+what+does+it+mean%3f&u=a1aHR0cHM6Ly93d3cuc3RhdGlzdGljc2hvd3RvLmNvbS9oZXRlcm9zY2VkYXN0aWNpdHktc2ltcGxlLWRlZmluaXRpb24tZXhhbXBsZXMv&ntb=1). [Heteroskedastic data can cause problems for some statistical methods that assume homoskedasticity or constant variance](https://www.bing.com/ck/a?!&&p=9f1babfea64b5130JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc5OQ&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+heteroskedasticity%2c+and+what+does+it+mean%3f&u=a1aHR0cHM6Ly93d3cuaW52ZXN0b3BlZGlhLmNvbS90ZXJtcy9oL2hldGVyb3NrZWRhc3RpY2l0eS5hc3A&ntb=1)[.](https://www.bing.com/ck/a?!&&p=b0f042b26db63e86JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTgwMA&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+heteroskedasticity%2c+and+what+does+it+mean%3f&u=a1aHR0cHM6Ly93d3cuaW52ZXN0b3BlZGlhLmNvbS90ZXJtcy9oL2hldGVyb3NrZWRhc3RpY2l0eS5hc3A&ntb=1)

12. Describe the concept of ridge regression.

Ans: Ridge [regression](https://www.mygreatlearning.com/blog/what-is-regression/) is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. It is hoped that the net effect will be to give estimates that are more reliable.

13. Describe the concept of lasso regression.

Ans: [Lasso(least absolute shrinkage and selection operator; also Lasso or LASSO) regression is a **method for stopping overfitting in linear regression models**](https://www.bing.com/ck/a?!&&p=d3c88cc71c1b33ecJmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc2Nw&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=Describe+the+concept+of+lasso+regression.&u=a1aHR0cHM6Ly93d3cudHV0b3JpYWxzcG9pbnQuY29tL3JpZGdlLWFuZC1sYXNzby1yZWdyZXNzaW9uLWV4cGxhaW5lZA&ntb=1).,is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.

[It uses **shrinkage**, which means that data values are shrunk towards a central point, like the mean](https://www.bing.com/ck/a?!&&p=7b2d06b0768f79f9JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc3MA&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=Describe+the+concept+of+lasso+regression.&u=a1aHR0cHM6Ly93d3cuc3RhdGlzdGljc2hvd3RvLmNvbS9sYXNzby1yZWdyZXNzaW9uLw&ntb=1). [It also includes a **penalty term** in the cost function, which adds the total of the absolute values of the coefficients rather than the sum of the squared coefficients](https://www.bing.com/ck/a?!&&p=6d596b473d9b87c4JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc3Mg&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=Describe+the+concept+of+lasso+regression.&u=a1aHR0cHM6Ly93d3cudHV0b3JpYWxzcG9pbnQuY29tL3JpZGdlLWFuZC1sYXNzby1yZWdyZXNzaW9uLWV4cGxhaW5lZA&ntb=1)[1](https://www.bing.com/ck/a?!&&p=7c1e02656920a982JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc3Mw&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=Describe+the+concept+of+lasso+regression.&u=a1aHR0cHM6Ly93d3cudHV0b3JpYWxzcG9pbnQuY29tL3JpZGdlLWFuZC1sYXNzby1yZWdyZXNzaW9uLWV4cGxhaW5lZA&ntb=1). [This encourages **simple, sparse models** with fewer parameters](https://www.bing.com/ck/a?!&&p=ded8e60918a1ea26JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc3NA&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=Describe+the+concept+of+lasso+regression.&u=a1aHR0cHM6Ly93d3cuc3RhdGlzdGljc2hvd3RvLmNvbS9sYXNzby1yZWdyZXNzaW9uLw&ntb=1). In statistics and machine learning, lasso (least absolute shrinkage and selection operator; also Lasso or LASSO)

14. What is polynomial regression and how does it work?

Ans: Polynomial Regression is a form of Linear regression known as a special case of Multiple linear regression which estimates the relationship as an nth degree polynomial. Polynomial Regression is sensitive to outliers so the presence of one or two outliers can also badly affect the performance.

15. Describe the basis function.

Ans: This is a generalization of linear regression that essentially replaces each input with a function of the input. (A linear basis function model that uses the identity function is just linear regression.)

16. Describe how logistic regression works.

Ans:  Logistic regression uses an equation as the representation, very much like linear regression. Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y). but it is used to make a prediction about a categorical variable versus a continuous one.